**OpenPress – Models Characterization**

Overview:

The goal of this document is to characterize the common models of the OpenPress Project. The models will be used both as common language between the different teams (Front-end, Back-End, OCR and any other team to join later), and as a reference for actual models to be implemented.

The Document contains the following parts:

1. **Objects Relationship Model** – Detailed listing of all Objects in the models.
2. **Objects Characterizations** – Attributes of all unique objects are detailed

## Reader’s guide:

Object Relationship Models contains three different models:

1. Physical Model – Describes the paper as a thing, with volumes containing pages.
2. Logical Model – Describes the paper as an ‘idea’, with volumes containing articles.

**Notes on those models:**

* Those two models are very similar, with the exception that, as mentioned, one contains a relationship of ..Section-Page-Paragraph, and the other a relationship of ...Section-Article-Paragraph.
* The relationship between Page and Article is of N..N.
* Unless otherwise stated, indentation in model means a relationship of 1..N
* **We have a few unanswered question that needs a decision before we can continue!**

1. Contextual Model(s) – Describes the world in which the press exists. Like people (people who write the press, or appear in the articles), significant time episodes, etc.

**Notes on those Models:**

* There’s more than one.
* They may be related to one another
* They must be related to at least one object in either one of the first two models.
* We detailed only those models we considered most relevant for this stage of the project, for obvious reasons.

# Objects Relationship Models

## Physical Model

Archive

Title (?)

Volume

Section

Page

Paragraph (Rectangle)

## Logical Model

Archive

Title (?)

Volume

Section

Article

Paragraph

* ?\* See question below.

Word

Photo

## Contextual Models

### Temporal Model

Calendar (Year, month,Day, weekofyear…hours)

Volume

Historical Episodes

(N..N)Volume

(N..N)Article (?) \* Manual? Extraction?

### Media Contributors Model

Person

(n..n)Title\_As\_Founder

(n..n) Volume\_As\_Chief\_Editor

\*(n..n) Volume\_As\_Contributing\_Editor – does exist separately from sections?

(n..n) Volume\_As\_Owner\_of\_Title\_at\_date\_of\_publication

(n..n) Section\_As\_Chief\_Editor\_of\_Section

(n..n) Section\_As\_Contributing\_Editor

(n..n) Article\_As\_Reporter

(n..n) Article\_As\_Photographer

\* Paragraph.photo/photo – depends on Implementation.

\* Are there any other stake-holders we would like to represent in the model (designers, donors, supporters, legal adviser, etc..)

Press Agencies

(n..n)Article\_As\_Source \* photo source or textual context

(n..n)\* Article\_As\_Secondary\_Source \* Do we have that kind of information?

### People

(n..n) Paragraph\_Automatically\_Extracted\_from\_Paragraph \* within textual object

(n..n) Article\_Manualy\_Marked\_as\_Mentioned\_in\_Article

(n..n) Article\_Manualy\_Marked\_as\_Related\_to\_Article

(n..n)Related to Historical Episodes

### Location

(n..n) Paragraph\_Automatically\_Extracted\_from\_Paragraph \* within textual object

(n..n) Article\_Manualy\_Marked\_as\_Mentioned\_in\_Article

(n..n) Article\_Manualy\_Marked\_as\_Related\_to\_Article

### Topic

(n..n) Paragraph\_Automatically\_Extracted\_from\_Paragraph \* within textual object

(n..n) Article\_Manualy\_Marked\_as\_Mentioned\_in\_Article

(n..n) Article\_Manualy\_Marked\_as\_Related\_to\_Article

\* Better define Topic and relations to content.

### Client-Activity

Article\_Watched

Searches\_commited

Marked \_as\_favourites

Shared\_Pargaraphs

Paragraph\_watched

\* What else? Do we want to keep history? Or session-long memory only?

# Objects Characterizations

**Attributes of all Unique Objects**

**Physical and Logical Models**

## Archive

ID

Name

Description

Archive Owner (Organization)

Archive Admin \*

(Number of Pages, Titles..)

(Date range)

## Title

ID \* What will ID Mean for us (e.g. delete and re add, do we want a hash that gives us information on the content?)

Name

Display\_Name

Description

Founder

Country of Publication

(Languages, current Language, language by date\_range)

(Number of Volumes, Page, Dates of publication..)

## Volume

ID

Number/Name

Standardized Number/name

Date of Publication

Staff

Chief Editor

Editorial Members \*

(Contributors)

Owner of Title at Time of Publication

Volume of Distribution \* Do we have it?

Language

Is\_weekend

Is\_unique\_publication\_by\_event \* What do we have?

(Sections)

(Number of pages, number of articles, number of photos, number of commercials)

## Section

ID

Name

Description

Section Editor

Section Staff\*

(Contributors)

(Dimensions- Width,height)

(Number of pages, number of articles, number of photos, number of commercials)

**\*\*\* Note – Till here Physical and Logical model were the same, not anymore! \*\*\***

## Logical.Article

ID

Article\_Title

Article\_Secondary\_Title

Contributor

Is\_commercial

## Physical.Page

ID

Page\_number

Dimensions – Width, Height, \* also padding, margins?

Is\_back\_cover

Is\_front\_cover \* Do we need this, or is it redundant?

(number of paragraphs, number of articles, contributors)

\* Tags – contextual model?

**\*\*\* Note – We’re back to objects that belong both to Physical and Logical models \*\*\***

## Paragraph

\*How should we treat the 3 possible parts of Paragraph – Title\_text, Content\_text and Photo? Are they all separated from each other, with 1..n relationship to Paragraph? Or is one of them an attribute of the other? If so, will it be 1..n or 1..1?

ID

Paragraph\_Number

Title

Representation\_By\_Photo - (\*Where the actual photo is, and it should be retrieved)

Photo \*If object exists separately in model

ID

Photographer

File/ Representation\_By\_Photo - (\*Where the actual photo is, and it should be retrieved)

\* Tagging of people, places.. – part of contextual model, but how is it applied actually?

## Content\_Text

ID

Text\_Representation\_By\_Photo

Text\_Representation\_By\_OCR (unicode) \* Will it be an ocr object with meta-data of its own? If so, we will also need concrete textual content as Unicode object, for easy manipulation.

## Title\_Text

ID

Text\_Representation\_By\_Photo

Text\_Representation\_By\_OCR (unicode) \* Will it be an ocr object with meta-data of its own? If so, we will also need concrete textual content as Unicode object, for easy manipulation.

**Contextual Models**

## Historical Episodes

ID

Name

Description

DateRange

## Person

ID

Name

..Biography..